

現代中国語における形容詞および その重畳形の分布マップの自己組織化

人間社会環境研究科 人間社会環境学専攻

林 智

Self-Organizing Map of Reduplicative Adjective in Modern Chinese

HAYASHI Tomo

Abstract

In modern Chinese there are many types of reduplications and its discussions by now.

This paper analyzes the distribution of reduplicative adjective in modern Chinese using statistic approach and self organizing map, the author evaluate and compare the distribution of adjective and reduplicative adjective in specified sentence structure, and detect that there are differences of the factors of their distributions between adjective and reduplicative adjective. Adjective distribution is related to their measure concepts and reduplications are not affected by it but related to another factors such as modification function.

Key Words

Modern Chinese, Adjective, Reduplication, Statistical analysis,

1. 形容詞と重畳

現代中国語において語の重畳は多くみられる現象であり、名詞・動詞・形容詞を基本単位として広く行われる。典型的には、例えば“大”という形容詞を“大大”と2回重ねて繰り返して用いる形式を指す。“清楚”に対して“清清楚楚”・“清清楚楚”、“糊涂”に対して“糊里糊涂”など原形が二音節の場合は重ね方が一層複雑になり、一般的にABB型、ABAB型、AABB型、A里AB型と分類されている。

この重畳に関しては、文法上・意味上・音韻上の様々な複雑な問題があり、研究も盛んに行われてきた。重畳に関する文法研究の先鞭は朱德熙、呂叔湘らがつけその研究の基礎を確立した。¹⁾ 朱德熙は重畳形の基本的な形式をまとめた上で、その差異から重畳形と原形との間の主な区別を行っ

た。また、呂叔湘はAA型、ABB型、AABB型、A里AB型などを含む広範囲の形式について、修飾語・述語・補語にどのように展開されるか詳細な研究を行い、現在における重畳形の様々な研究の基となっている。

重畳形式の研究は、その文法構造上の問題、たとえば、AABB型の重畳形式において、その由来がABという原形の重畳からAABBという形式が生まれたのか、或いはAAという重畳形式とBBという重畳形式が発生し、両者が合成する事によってAABBという形式が生まれたのか、といった内部構造にまつわる問題やA里AB型の組成に関する問題など様々な側面について議論が行われている。²⁾

また、文法機能上の問題としては、赵元任や呂叔湘などが前置される副詞との関係や連体修飾を行う時と連用修飾を行う時の後置詞の差異を論じ

ており、³⁾ 朱德熙は形容詞の重畳形が連体修飾する際の「名詞化」、すなわち後方に“的”を伴って名詞を修飾する時に体言と化していることを方言における重畳形と的の振る舞いを論拠として証明し、形容詞の重畳形が述語の位置に来る場合は用言としての機能を持つとした。さらに、重畳形の原形が性質形容詞であり重畳式が状態形容詞であるため、前者が生起上の制限をより多くうけることをあげて、その差異を明らかにしている。⁴⁾

形容詞の重畳形の意味的な議論については、重畳する事により原形からどのように文法上の意味特徴を変えているのか、という点において異なる面での解釈が行われている。一つの側面としては、「程度を表している」、すなわち、重畳する事でその数量や尺度といった量を増大（あるいは減少）させているという考えがある。⁵⁾ “大大”（とても大きい）は“大”（大きい）の大きさの量を増加させているのであるという解釈である。別の側面として、「状態を表現」しており、重畳する事により状態変化の中において表現上のベクトル（方向感）を持たせることにより、聞き手によりリアルな情感を与え、かつ発話者の肯定感や感情上の方向性を描写する機能を果たしているという主張もある。また、重畳形が「程度」の大小を表しているのではなく「強調」を行っているという側面も論じられている。⁶⁾

このような最も基本的な議論に加えて、中国語の重畳形に対応する他国語の翻訳語や重畳形との対応や、重畳形式を1つに絞って詳細に論じた議論など、形容詞以外の議論も含めて、中国語における重畳形に関する問題は数多くある。

本稿では、この現代中国語における形容詞の重畳形について、過去に行った単音節形容詞についての分析を踏まえた上で（3-1節）、統計手法を用いてその分布状況を調査し、その上で重畳形形容詞の分布上の偏向が原型である単音節形容詞と同じような偏向をもって表れているのかを明らかにする。また、複数の構文形式についての分布状況を多変量解析を用いて、形容詞の重畳形のク

ラストリング（特徴ベクトルによるカテゴリー化）を試み、複数の構文形式における単音節形容詞のクラスタリングと比較することで、原形・重畳形双方の形式の大局的な生起分布に中核的な影響を与えている文法機能や語義、すなわちクラスター間において差異が大きく表れる語義・特徴が、原形から重畳される事によりどう変化するかを明らかにする事により、原形と重畳形が異なるサブカテゴリーを示す事を証明する。

2. 統計の方法

統計の具体的な方法であるが、形容詞の重畳形を計量的な側面から分析するために、現代中国語の書き言葉をおさめた品詞タグ付きコーパスに対する基本的な頻率分析、さらには多変量解析の一種である自己組織化マップ（Self-Organization Map）による総体的な解析の両者の手段を用いる。⁷⁾

かつて論じた形容詞の分布に関する論文の中でもすでに触れた事であるが、コーパスと統計を用いた計量言語学の手法でもって、実際の発話環境（書き言葉に限定されてはいるが）に対する分析を行う事で、語感のみに頼るだけではなく、数量というある意味客観的な論拠と、真か偽か言い換えれば0か1かという2値ではなくその中間の段階を表現しうる方法と、さらに複数の特徴（属性・パラメーター・要素間の分布上の差異）を同時に俯瞰するという視点を獲得することができる。⁸⁾

本稿で用いるコーパスは、国家語委現代汉语通用平衡語料庫（CYLK）であり、これは様々な分野のテキストに対してメタ情報と精度の高い品詞タグ（POS タグ）を付加したものであり、これを分析用にXML形式に変換したものである。

分析環境に関しては、統計のパフォーマンスを考慮し、林（2005）において構築したネットワーク上のデータベースサーバにコーパスのデータベースを設置しクライアントから検索クエリを行った形ではなく、ローカル環境のメモリ上に展開したファイルベースのデータベースを用いる事により、従来の数十倍の速度でクエリを行えるように

改良したものをを用いている。⁹⁾

3. 単音節計量形容詞と重畳形

まずは単音節形容詞の分布状況とその重畳形の分布について解析を加える。単音節形容詞は計量形容詞と属性形容詞に大別でき、前者には量的な概念が存在する。以前の論文でも述べたように、単音節形容詞の計量性とその大小の方向性は構文における生起分布に対して大きな影響を与える事がはっきりしている。計量性がある場合はその尺度の大小が影響を与え、ペアの中で尺度の小さい方(「大きい小さい」の場合は「小さい」、「高い低い」の場合は「低い」)が生起に制限を受けやすい。具体例をあげると、“X有Y那么A”(XはYのようにAだ)という構文のAには形容詞が生起できるのであるが、「小さい」「低い」などの形容詞はほぼ生起していない。また、発話者の心理的な良い評価、悪い評価を表す褒貶義(ほうへんぎ)も同じくそのプラスとマイナスの方向性が分布に影響する事が分かっている。属性形容詞にはこのような尺度による生起率の変化は大きくは見られず、発話時に褒貶義が加わる事により若干の生起率の変動がみられるくらいである。ただし、これらの生起上の制限については時代によって少し制限の緩まりの傾向を見せているらしいが、完全にその傾向にあるとははっきり言えるほどまでではない。¹⁰⁾

これらの分布上の偏りは単音節形容詞の持つ属性の違い、すなわち形容詞グループ内部において、計量性の有無やその尺度上の大小、褒貶義の存在が中国語形容詞のサブカテゴリーを構成する主要な要素である事をはっきりと計量的観点から示しているが、重畳形になった形容詞でははたして明らかな偏りが見られるのかどうか解析を行う事にする。

まずは、代表的な、つまり生起率の高い単音節形容詞とその重畳形についてであるが、これについては、原形と重畳形の生起率を調べその比率を

見る事でその単音節形容詞がどの程度重畳形になりやすいか、という尺度を得る事が出来る。もし全ての原形となる形容詞が偏りなく重畳形に移行できるのであれば、比率自体の大小はともかく全てのペアにおいて生起率の比率には大きな違いはみられないはずである。もし比率に大きな違いがあるのであれば、重畳形への移行には偏りがあり、そこに何らかの原因が存在する事を表している。

上に述べたとおり計量形容詞の程度の大きい方と小さい方にはもともと生起率に違いが大きい、重畳形になる事で尺度の増加を志向するのであれば、ペアの両者に生起比率の大きな違いがみられるはずである。もし差が少なければ、重畳することによって尺度の増加だけを志向しているのではなく、増加・減少のいずれの方向にも等しく志向できる事になる。つまり、重畳形は原形の語義をより強調する意図でしばしば使われるのであるが、より大きく、高く、重く、といった増加方向への志向性と、より小さく、低く、軽くといった減少方向への志向性に、原形と同じく重畳形においても差異があるかどうかははっきりする。

ここでは、“高低”“大小”“多少”“长短”“重軽”という計量形容詞のペアについて、上記の統計をとる。これらの形容詞は生起数が少なくとも1500以上はあり、比較に十分なサンプル数を確保していると考えられる。

まずは、“高低”の分析であるが、“高”の生起頻率に対する“高高”の生起頻率の比率は0.0407であり、“低”の生起頻率に対する“低低”の生起頻率は0.0231である。これはおよそ2倍弱の差であり、“高”に比べて“低”が重畳形になりにくい事を表している。この結果は比較構文などにみられた計量性のベクトルによる生起制限のように、計量形容詞ペアの程度の小さいものは生起しづらい現象が生起条件に影響している可能性を示唆している。

次に、“大小”であるが、“大”の生起頻率に対する“大大”の生起頻率の比率は0.0652であり、“小”の生起頻率に対する“小小”の生起頻率は0.05240である。これは両者にほとんど差がない

事を表しており、計量性のベクトルによる生起の制限はほとんどかかっていない事を表している。

次に、“多少”については、“多”の生起頻率に対する“多多”の生起頻率の比率は0.0063であり、“少”の生起頻率に対する“少少”の生起頻率は0.0018である。この結果が表しているのは、“多少”ペアはそもそも上の2者よりはるかに重畳形になりにくい形容詞であり、また、単音節の時のように“多”の方が3倍程度重畳形になりやすいという事である。

次のペアの“长短”は、単音節の場合、特定構文において典型的な生起の違いを表していた形容詞であるが、これらの差異はどうなっているかというと、“長”の生起頻率に対する“长长”の生起頻率の比率は0.0850であり、“短”の生起頻率に対する“短短”の生起頻率は0.1609となっている。両者とも全てのペアの中でも高い数値を表しているのであるが、頻率は逆転して、尺度の低い“短”の方が比率を見せている。

最後のペアである“重軽”についても、同じような傾向で、“重”の生起頻率に対する“重重”の生起頻率の比率は0.1656であり今までの形容詞ペアの尺度の大きい方の形容詞の中で最も高い数値であるものの、“軽”の生起頻率に対する“轻轻”の生起頻率は0.8であり、重畳形が原形の生起頻率にせまっており、原形として使われる頻度と重畳形として使われる頻度が相当近く、他の形容詞と比べても突出して高い事が分かる。

以上の結果から、次の事が推定できる。第一に、計量形容詞ペアの重畳形については、“大小”を除いて両者に2倍程度の比率の差があり、これは両者の生起条件に何らかのベクトルがかかっている事がわかる。ただし、その理由は明らかに計量性のベクトルがプラス方向か、マイナス方向か、という原型における生起条件の差異の原因とは異なるものであるという事がわかる。第二に、同じ計量形容詞ペアでありながら、“高低”“大小”“多少”と“长短”“重軽”では生起自体の確率とどちらがより重畳形になりやすいか、という点につ

いて正反対の特徴を見せているという事である。

それでは、この原因はどこからきているのかさらに進めて分析を加えると、これらの形容詞が文法上どういう機能を果たしているか、という点で大きな違いがみられる事がわかった。それはすなわち、あらわれた重畳形全体に対して副詞として用いられている度合いを見れば明らかに違いがわかるのであるが、前3者のペア（“高低”“大小”“多少”）において副詞として表れている比率と、後2者のペア（“长短”“重軽”）において副詞として表れている比率を比べると、後者の比率の方が高く、つまり、副詞としての定着の度合いによって、後2者のペアは遙かに重畳形になりやすく、また、その中でも程度の低いマイナス方向のペアが重畳形になりやすいという結果を表していると考えられる。

以上のように、計量形容詞のペアについては、重畳形へのなりやすさ、或いは重畳形の定着性という観点からいえば、計量形容詞がもっていた尺度、ベクトルの方向性というのは、分布上既に重要な要素ではなくなっており、むしろ、特に副詞の機能がどれだけ使われやすいか、といった要素がより重要な課題になっている事がわかる。これはすなわち、1. で述べた重畳形の意味的な議論の中で、「程度」、「状態」、「強調」といった要素のうち、「状態」、「強調」と区分されている要素が分布状況に一番影響を与えやすい事を示唆しているのではないかと考える。もし、重畳形の定着性の差異に「程度」の側面、つまり尺度の大小の深化という機能、あるいはその大小の方向性が重大な影響を与えているのであれば、分布への影響もはっきり表れているはずにもかかわらず少なくともこの結果からは全くそれが見られない。しかし、さらに厳密にそれを証明するためには、品詞タグが機能によってもう少し微細にカテゴリー化されていないと難しい。

4. 構文との関係の多変量解析

重畳形形容詞が生起している大まかな状況であるが、重畳している AA という形式の前後の品詞パターンから、ほとんど影響のないと思われる大量のコーパス中から少数しか出ないパターンを除くと、ほぼ以下の形式に集約できる。

・ AA 地 (的) V

你长长地吐一口烟。(陈朝声『星海』)
「(あなたは) 長々と煙草を吸った。」
一定要吃三顿饭, 要多多的吃, (董向荣『额木尔脱险记』)
「必ず3杯食べなさい、たくさん食べなければなりません」

・ AAV

彭莉当上模特后, 穿衣服的开支大大增加, (华培明『走向世界的中国时装模特』)
「彭莉はモデルになった後、服の支払いがどんどん増え」

・ V 得 AA 的

牛角似的大勾嘴伸得长长的, (王世阁『在长白山森林里』)
「牛の角のような大きな勾嘴(魚)は伸びて長く長くなり」

・ AA 的 N

长长的毛拖到了地上, (曹桂林『北京人在纽约』)
「長々とした毛皮は地面に落ちて」

・ AA+N

到达高高云天上的时候, (赵淑侠『我们的歌』)
「高い高い雲の上に出た時」

・ AA (的),

体委张世幸吹起了集合哨, 哨音长长的。(邵英、子实『万岁, 高3・2』)
「体育委員会の張世幸は集合の笛を鳴らし、笛の音は長々と鳴り響いた」

これら的大別して6つの構文形式が、コーパス内において重畳形形容詞とよく生起するパターンであり、これ以下の解析ではこの6つの構文への形容詞ごとの生起のしやすさから解析を行うことで、その差異と傾向を明らかにするものである。

解析の方式は以前と同様で、重畳形の形容詞が表れる全てのパターン(構文形式)内におけるそれぞれのパターンの比率を統計から計算し、その比率を特徴ベクトルとして多変量解析を行うことにする。

本稿で用いている多変量解析・クラスタリングの手法についてはすでにいくどか述べているので、多くは述べてないが、6つの構文における生起頻率を6次元の特徴ベクトルと見た場合に、教師なし学習を行うニューラルネットワークを構築する事によって、近い特徴ベクトルを持つノード同士をまとめていく事で、最終的に6次元空間上に展開する特徴点(これが対象の重畳形形容詞が特徴空間上で存在する点なのであるが)を2次元上のマップに投射する事になる。このマップ上でのクラスターが対象となる重畳形形容詞の分類を表す事になる。

対象となる重畳形形容詞は、以前の解析と同じく生起頻率の高いものから順に選んでいる。

下に解析の結果となるマップを示す。SOMに

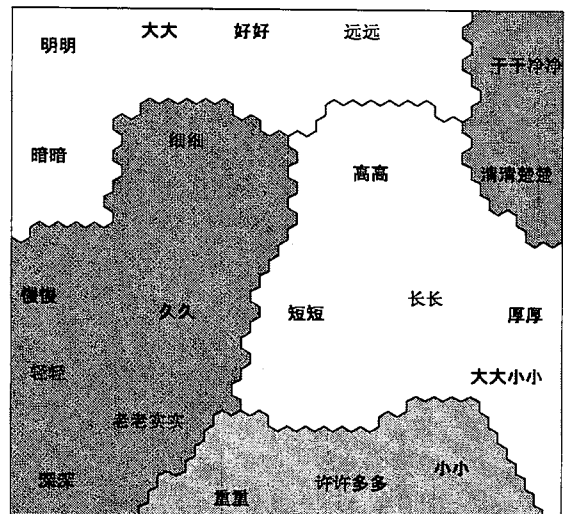


図1 全体の SOM

においてはマップ上のクラスター間の距離自体は直接ベクトルの距離を表しているのではなく、その位置関係が特徴ベクトルの近接関係を表している。

マップ上で、重畳形容詞は大きく5つのクラスターに分類されている。“干干净净/清清楚楚”2語で独立したクラスターを構成しており、後は“大大/好好/远远/明明/黑黑”が上部のクラスター、左側に“轻轻/慢慢/久久/细细/深深/老老实实”のクラスター、中央から右に“高高/短短/长长/厚厚/大大小小”、下部に“重重/小小/许许多多”というクラスターの分類になっている。

初めにあげた右上のクラスターは、二音節形容詞の最もよくつかわれ、定着性も高いと思われる2つの単語がまとまっており、特徴がよく似ているという点に不思議はない。残りのクラスターについ

ては、もう少し精細な分析が必要となろう。ここで、6つ特徴ベクトルがどのように分布しているのか、ヒートマップを用いて観察する事で、それぞれのクラスターの特徴がさらに明らかになるので見ていくことにする。ヒートマップは明暗が頻率の大小を表している。(明るい方が大)

まず、下部のクラスターを特徴づけているのは、AA+Nの形式が非常に多く、ついでAA的Nが高めであり、AA+Vなどは低めであるという点である事がわかる。これは、“重重/小小/许许多多”が比較的名詞と結びつきやすい事を表している。

逆に、動詞と結びつきやすいのが左の“轻轻/慢慢/久久/细细/深深/老老实实”となっており、このクラスターは総じて連体修飾形にはなりにくい。上部の“大大/好好/远远/明明/黑黑”というクラ

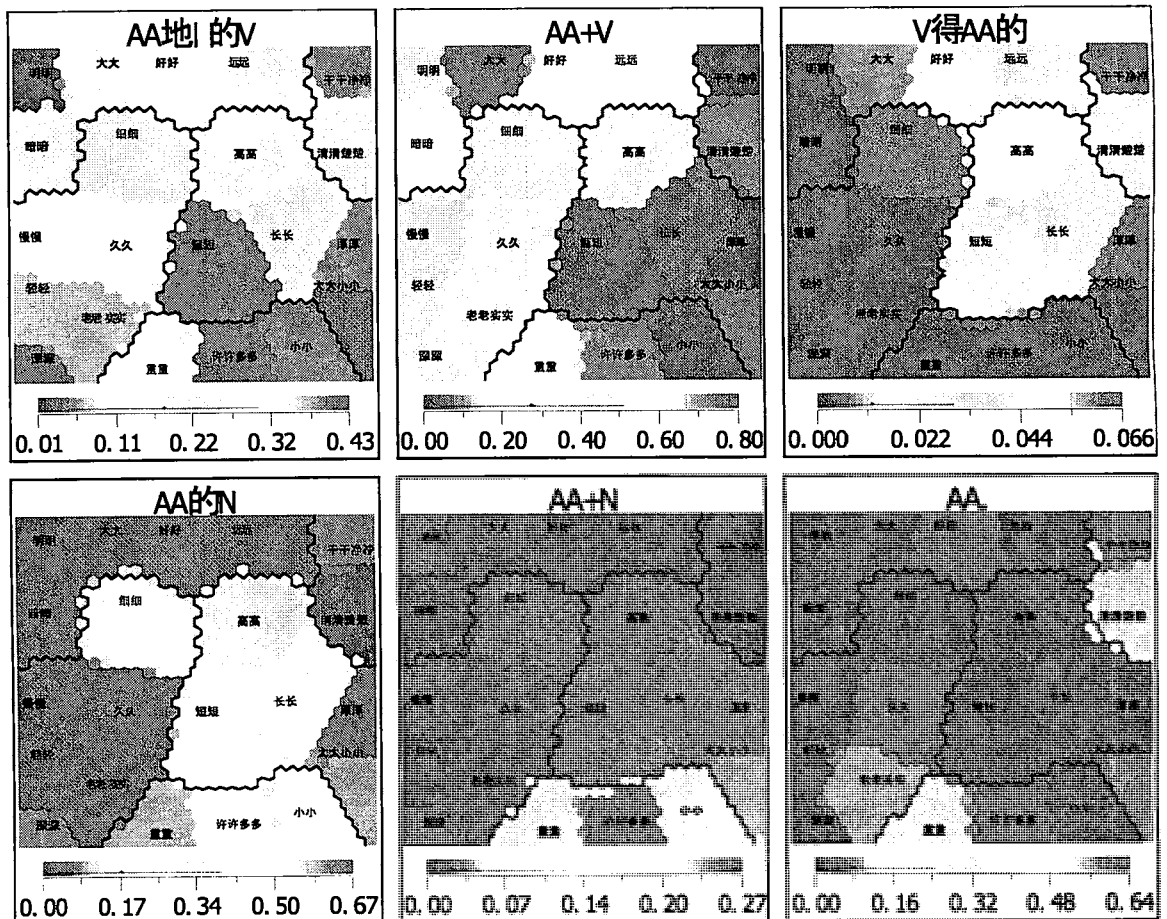


図2 ヒートマップ

スタも同じく動詞と結びつきやすく名詞と結びつきにくいことを表している。この両者の違いは、連用修飾するとき、“地/的”を伴う頻度が異なる点である。

中央のクラスタである“高高/短短/长长/厚厚/大大小小”はAA的Nという用法で頻繁に使われやすいのが第一の特徴となっている。

このように、重畳形形容詞の構文上の分布状況は、連体修飾・連用修飾との分布状況の違いと“的/地”と共起するかどうかという2点で大きく特徴づけられており、これらの差異から5つほどのグループに重畳形形容詞が分けられる事がわかる。また、グルーピングされた形容詞の原形の特徴、すなわち計量性や尺度となるベクトルの方向性(プラス・マイナス)などの原形が持っていた語義は、これら重畳形の構文の分布状況と明らかに関係がなくなってしまう事がはっきりと示されている。もし、原形のように計量性が重大な影響を与えているのであれば、原形の分布状況と同じように計量性の大小によってまずグルーピングが行われるはずなのにも関わらず、上記に示したようにここではまずその傾向はみられないからである。

5. おわりに

以上にのべてきたように、形容詞の重畳形は計量的な観点からいって原形となる形容詞が備えていた語義・特徴と分布の関連性が全く異なり、全く違った体系によるものである事を明らかにした。原形の分布においては、計量性の有無や尺度の大小が分布にもっとも大きな影響を持っており、次いで褒貶義の有無や単音節か二音節かという特徴が影響を与えているのに対して、重畳形形容詞においては連体修飾・連用修飾のどちらとして用いられるかといった側面が大きな影響を与えている事が分かった。すなわち、重畳形として原形よりも程度の大きさ・小ささが一層深い事を表す表現や、発話者の強調や感情を表す表現は原形と異な

り計量の方向性に生起条件を左右されない。一つの原因として考えられるのは、特に原形において生起条件に偏りが大きい比較構文に生起するか否かという点が分布の要因の重要性に影響を与えていると考えうる。

現状のコーパスから得られる情報からは、文の文法機能、文脈情報を得られる事ができないため、この体系を構成している要素がどのようなものであるかを計量的な手段ではっきりと明かす事はできないが、近年整備されつつあるさらに文法情報が付加されたコーパスをもちいれば、原形・重畳形形容詞のクラスタを構成する要素に文脈や文法機能の情報を付加でき、今は構文の形式のみで等しく取り扱わざるを得ないためそぎ落とされているクラスタを構成する特徴を抽出でき、さらに深い解析が可能になるであろう。これまでの解析では検索の多大な試行と頻率の大きい現象に絞る事によって形容詞群、構文群の抽出を行い解析を物理的、時間的に可能な範囲にまでかなり縮めているが、文法タグが付加されたコーパスを用いればまずその必要がなくなり、また特徴の次元を新たに増やす事ができることで、現在の特徴次元では見えていない分布上の特徴を明らかにする事ができる。そのうえで重畳形形容詞の振る舞いについて一層の理解を深め解釈をはっきりさせる事が当面の課題である。

文献目録

- T.Kohonen, (1999). Self-Organizing Maps. Springer.
 吉田清香, 林智, (2001). 同程度を表す2つの比較構文とその差異の考察. 金沢大学中国語学中国文学教室紀要, 第5輯, 31-62.
 朱德熙, (1956). 现代汉语形容词的研究. 语言研究 (1), 83-104.
 朱德熙, (1979). 语法讲义. 商务印书馆.
 石毓智, (1986). 论汉语的取法重轻. 语法研究 (2), 1-11.
 林智, (2005). 中国語言語コーパスを用いた研究とソフトウェア環境. 金沢大学中国語学中国文学教室紀要, 第8輯, 23-47.
 林智, (2007). 特定の構文グループにおける形容詞の分布: 現代中国語大規模コーパスによる統計とクラスタリング. 金沢大学中国語学中国文学教室紀要, 第10輯.

- 吕叔湘. (1965). 形容词使用情况的一个考察. 中国语言 (6), 419-431.
- 吕叔湘. (1980). 现代汉语八百词. 商务印书馆.
- 张敏. (2001). 汉语方言重叠式语义模式的研究. 中国语文研究, 24-39.
- 赵元任. (1990). 汉语口语语法. 商务印书馆.

[注]

- 1) 朱(1956)を参照。
- 2) 吕(1965)を参照。
- 3) 吕(1980)を参照。
- 4) 朱(1979)において“的”の用法を3種に分け、方言における使い分けと併せて詳しく論じている。
- 5) 石(1986)を参照。
- 6) 朱(1979)を参照。
- 7) Kohonen(1999)がその基本的な概念を説明している。
- 8) 林(2007)を参照。
- 9) 林(2005)において基礎的な環境を確立している。
- 10) 吉田・林(2001)において基礎的な分析を行い、林(2007)において統計的な解析を加えている。